

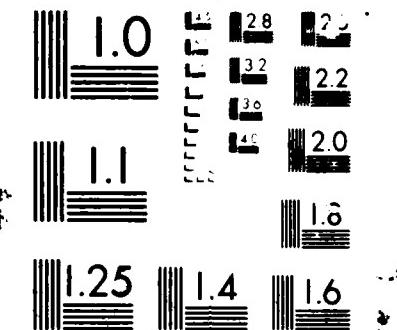
AD-A177 938 DNT: SPEECH COMMUNICATION(U) MASSACHUSETTS INST OF TECH 1/1
CAMBRIDGE DEPT OF ELECTRICAL ENGINEERING AND COMPUTER
SCIENCE L A WHEELER 27 FEB 87 N00014-83-D-0689

UNCLASSIFIED

F/G 5/7

ML







DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139

17

AD-A177 938

February 27, 1987

Dr. Leslie Anne Wheeler
Room 36-511Mr. E.G. Gouge
Project Manager
ASEE Suite 200
Eleven Dupont Circle
Washington, D.C. 20036Contract N00014-83-D-0689
Title: ONT Final Technical Report;
Speech Communication

Dear Mr. Gouge,

Here is my final technical report for my postdoctoral work supported by the ONT-ASEE postdoctoral fellowship program. It gives an overview of my activities during the past year and a description of the research I have been doing here at M.I.T. in Professor Kenneth N. Stevens Speech Communication laboratory.

I apologize for the delay in submitting a final report, but this has been an especially hectic time. I began working on this project only last September-October, and it is not the sort of project that can be concluded in a few months. The data are rolling in faster than I can reduce them, and the terminal is now swearing at me in four languages, three human and one machine. So, what I am sending is a kind of cross between a progress report and a proposal, which, to be frank, I wrote only to get my last paycheck, but it is proving useful in other ways.

This past year has been eventful, to say the least. Though my morale went up and down like a sine wave during the uprootings and upheavals, I have found my period of tenure to be highly rewarding, both educationally and scientifically. I am very glad that this fellowship program was created, and I would like to thank both the Office of Naval Technology and the American Society for Engineering Education for allowing me to participate in it.

Me ka laule'a,

Leslie Anne Wheeler, D.Sc.

MAR 11 1987

OIE FILE COPY

87 3

0 045

TECHNICAL REPORT

ONT-ASEE Postdoctoral Fellowship

Dr. Leslie A. Wheeler

Overview:

My one-year postdoctoral tenure was roughly divided into three parts:

- The first two months were spent at Naval Ocean Systems Center in Kailua, Hawaii. While waiting for an interim security clearance, I spent my time
(1) writing and revising scientific articles for myself and others,
(2) designing and refining new research projects to be carried out in collaboration with NOSC scientists,
(3) gaining familiarity with NOSC microcomputers by programming an original pattern recognition algorithm from my doctoral work.

- When my interim clearance was refused, I was asked to leave the base and was told that my stipend would not be renewed. During the next few months, I
(1) attended a scientific meeting in Cleveland and was accepted into the laboratory of Professor Kenneth Stevens at Massachusetts Institute of Technology (Research Laboratory of Electronics), as of August-September, for work on speech communication,
(2) made professional contacts with scientists studying dolphin psychoacoustics on the West Coast,
(3) returned to Hawaii to study Hawaiian language and culture,
(4) again touched professional base in California,
(5) attended two more scientific meetings in Canada,
(6) moved to Cambridge, MA.

I paid all expenses for these activities out of my stipend.

- During the third part of my tenure, I
(1) presented a paper at a scientific meeting in Anaheim, CA (Acoustical Society of America),
(2) worked at M.I.T. on a cross-language study of nasalized speech sounds.

The following report concerns only my M.I.T. work: what I have done up to now, what I am presently doing, and what I intend to do.

Accreditation

NTIS Serial

STIC TAB

Uncontrolled

Justification

Distribution

Availability

Part 1



Background:

All human languages show certain common properties which reflect the underlying common genetic coding for acoustic communication in the human species. The semantic code of a spoken language consists of perceptual units of individual sound, the smallest of which are presently considered to be vowels and consonants⁽¹⁾. These units are organized in hierarchical fashion into syllables, words and sentences. The production and perception of the speech code are both based on certain articulatory shapes and movements of the vocal tract.^(1,2).

One property of language which is now considered universal is the use of nasalized sounds interspersed with oral sounds to transmit the speech code.⁽³⁾ The number of nasalized sounds important to this speech code, and the degree to which they are nasalized, may vary greatly from language to language. At one extreme end of the nasal functioning continuum is a language called "Puget Sound Salish", in which only one word of adult language has a nasalized segment. At the other end of the continuum is a language spoken in Paraguay called "Guarani", which has such an impressive collection of nasals that it also has a kind of nasality accent by which all important speech sounds are characterized as nasal or non-nasal⁽⁴⁾. In some languages, such as Hindi or Gujarati⁽⁵⁾, some oral vowels have nasal equivalents, such that the presence or absence of nasalization determines the meaning of a word. In other languages, such as English or present-day Hawaiian^(6,7), this distinction is considered not to occur. The introduction of nasalization to their otherwise oral vowels seems to have no apparent linguistic function, though it may actually have psychoacoustic importance, enhancing or diminishing intelligibility according to when or where it occurs during speech⁽⁸⁾.

One of the ongoing projects of Professor Kenneth Stevens and his colleagues at M.I.T. is a cross-language study of nasal or nasalized vowels in relation to corresponding non-nasal vowels. Their study has two purposes: (a) to investigate the timing of the sequence of articulatory events that occur when a nasal vowel is produced in the environment of certain types of consonants, and, (b) to determine for nasal or nasalized vowels the acoustic properties that could contribute to a universal nasal vs. non-nasal distinction for vowels⁽⁹⁾.

Professor Stevens and his colleagues have approached the problem in three ways. One involves acoustic analysis of nasal, nasalized, and non-nasal vowels within natural or nonsense utterances. Another involves modeling the articulatory events that occur during vowel production. A third involves perceptual studies of synthetic utterances which are modified in various ways to introduce various degrees of nasalization^(5,9). The results suggest that listeners from different language backgrounds base their identification of the nasal-non-nasal distinction on similar criteria. There seems to be a basic acoustic property of nasality, independent of the vowel, to which the auditory system responds in a distinctive way regardless of language background. However, there also seems to be one, or more, additional acoustic properties that could be used to various degrees in different languages to enhance the contrast between a nasal vowel and its non-nasal counterpart^(9,10).

The results of these studies raise many questions. What precisely are the acoustic features of nasalization which are common to all human languages? What are the language-specific variations on these universal themes? Why are they language-specific? What are the articulatory and perceptual limits on these variations? What can be done to acoustic properties to improve, for each language, the quality of synthetic speech?

The languages or language backgrounds studied so far by Professor Stevens and his group are Hindi, Gujarati, Bengali, Portuguese, English, and French. French listeners in the perceptual study disagreed with those of other language backgrounds on the degree and quality of nasalization needed to render various portions of the synthesized utterances natural in their language. They reported that the initial portion was not nasalized enough, while the final highly nasalized portion was much too prominent psychoacoustically (9). Because I am bilingual English-French, it was felt that I could contribute to the study by designing a French language data base which would allow investigating the properties of French nasal vowels and then determining the way in which they differed from analogous portions of Portuguese and English nasalized vowels.

Data base:

There are presently four nasal vowels in French. In theory, there are four corresponding oral vowels. In practice, however, factors such as context, usage, and stress and intonation erode their true correspondence, such that any remaining equivalences are exceptions that suggest that a general rule did once exist.

My data base contains all four nasal vowels. Each is situated within a "natural utterance", i.e., a word, or combination of words, from the French language. Each utterance is, in turn, situated within a carrier phrase:

"Dites ' ____ ' pour moi."
("Say ' ____ ' for me .")

Twenty-four of the 31 words were chosen such that each nasal vowel is followed by one of six consonants having a similar manner of articulation (stop consonants). These consonants can be grouped according to their place of articulation (bilabial: b,p; alveolar: d,t; velar: g,k) or to their voicing features (voiced: b,d,g; unvoiced: p,t,k). In two additional utterances, a nasal vowel is followed by a nasal consonant /n/. The remaining five utterances consist of isolated nasal vowels repeated three times, four of which show phonemic (cognitively meaningful) distinctions, and the fifth is a nasal vowel which once was linguistically distinct but is now considered to be indistinguishable from one of the other four (See Table 1).

Methods and results:

σ⁷

Three native French speakers were asked to read the list of carrier phrases three times. Recording was made on a Nakamichi LX-5 discrete head cassette recorder with an ALTEC 684-A microphone.

Analog-to-digital and digital-to-analog conversions were performed on a VAX-750 computer using an llk laboratory Peripheral Accelerator outfitted

Table 1.

1. Dites "mon oncle" pour moi.
2. Dites "un blanc" pour moi.
3. Dites "compas" pour moi.
4. Dites "un un un" pour moi.
5. Dites "incrédule" pour moi.
6. Dites "en cage" pour moi.
7. Dites "continent" pour moi.
8. Dites "imputable" pour moi.
9. Dites "on on on" pour moi.
10. Dites "un goût" pour moi.
11. Dites "en pierre" pour moi.
12. Dites "condiment" pour moi.
13. Dites "indignité" pour moi.
14. Dites "en dette" pour moi.
15. Dites "an an an" pour moi.
16. Dites "un coup" pour moi.
17. Dites "intimité" pour moi.
18. Dites "en gage" pour moi.
19. Dites "combat" pour moi.
20. Dites "in in in" pour moi.
21. Dites "ingrédient" pour moi.
22. Dites "un toit" pour moi.
23. Dites "en tête" pour moi.
24. Dites "congrès" pour moi.
25. Dites "en en en" pour moi.
26. Dites "un plan" pour moi.
27. Dites "imbuvable" pour moi.
28. Dites "un doigt" pour moi.
29. Dites "concret" pour moi.
30. Dites "en bière" pour moi.
31. Dites "mon ongle" pour moi.

with a KW-11K programmable real-time clock, an AD-11K 8-channel (differential) 12-bit \pm 5 volt a/d converter and an AA-11K 4-channel 12-bit \pm 5 volt d/a converter. The audio input device was a Yamaha Model K-1000 cassette tape recorder.

The digitized waveforms were edited and then analyzed in various ways using software written by Professor Dennis Klatt or Rich Goldhor. The displays illustrating this text were obtained using the programs RECORD (amplitude spectra and expanded waveforms), KLSPEC (spectral analysis performed by windowing a portion of the waveform, computing the discrete Fourier transform and weighting the summed squared dft values to form filters) and SPECTO (spectra computed every 5 ms from the input waveform files). We are beginning to test our results through formant and harmonic speech synthesis (KLSYN and HARSYN).

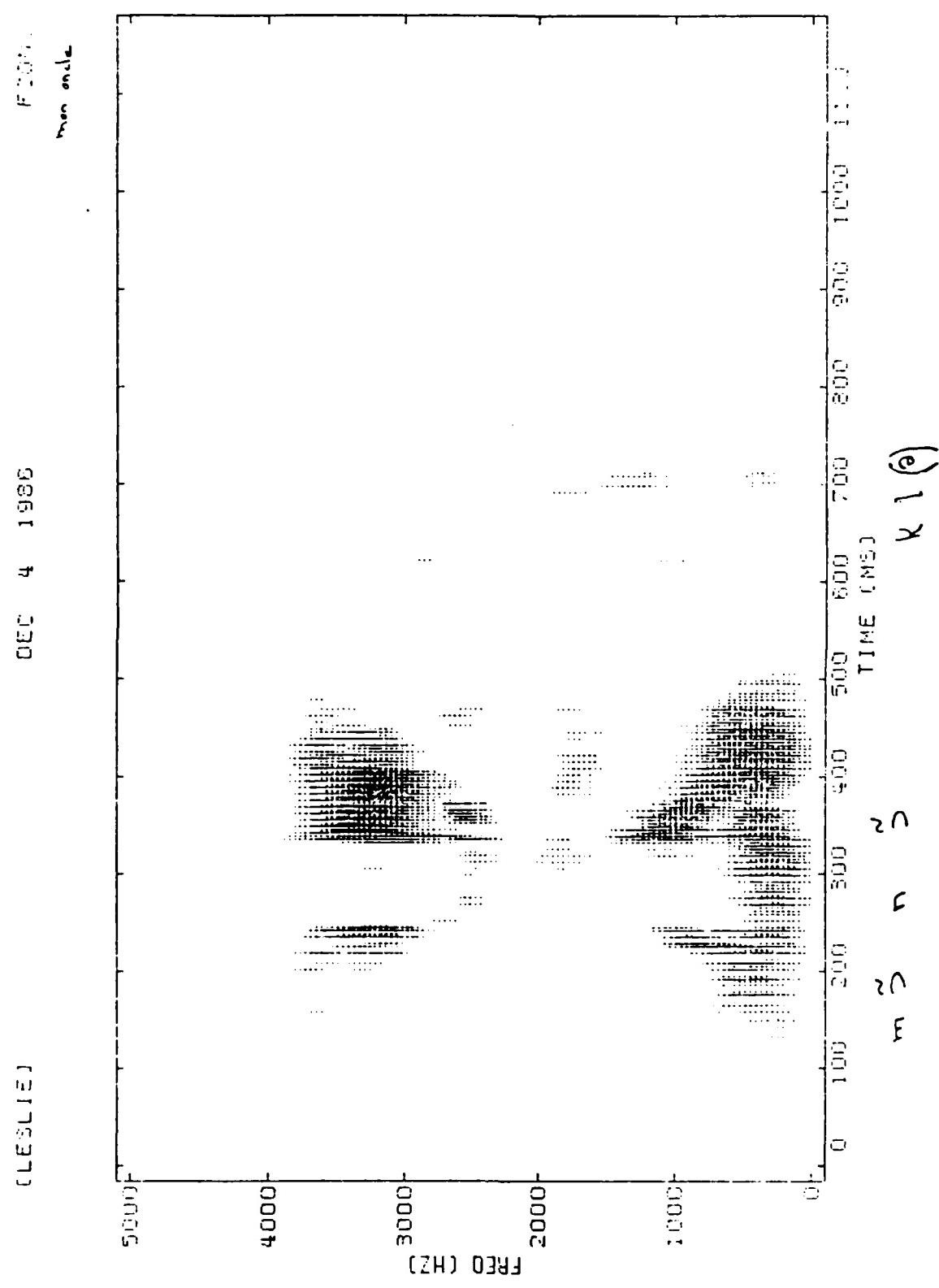
We first looked at durational characteristics of the different vowel segments, particularly the final portion which we call the "nasal murmur". The term is controversial when applied to French nasal vowels. It is actually a result of the oral tract closure and coupling of the oral tract with the nasal passages and sinuses which occur for production of a nasal consonant such as /n/ or /m/ or /ŋ/. Because the movement downward of the velum for a nasal consonant begins well before the beginning of oral tract movement toward occlusion, parts of vowels preceding and following nasal consonants are nasalized in all languages studied. Despite the way in which nasal vowels appear in the written language ($\acute{e} \rightarrow$ in; $\grave{a} \rightarrow$ en; $\grave{e} \rightarrow$ on; $\grave{u} \rightarrow$ un), there is no true nasal consonant preceding a stop consonant ($k\acute{e}$ dim \grave{a} \rightarrow condiment). We feel justified in calling the final portion of a French nasal vowel a nasal murmur, when it immediately precedes a stop consonant because its spectral properties resemble closely those of nasal consonants.

Figures 1, 2, and 3 are spectrographs of three French utterances in the study: "mon oncle", "condiment", and "continent", respectively. Figure 4A illustrates the transition from the vowel / \acute{e} / into the nasal consonant /n/ within the words "mon oncle", where the nasal consonant is actually pronounced. Figure 4b illustrates the transition from the same vowel into the stop consonant /d/ within the word "condiment". The nasal consonant is "effaced", i.e., it does not exist linguistically. In Fig. 4C, the corresponding transition is shown for the word "continent". Fig. 5 shows spectral sections of the three nasal murmurs (A: mon oncle; B: condiment; C: continent). The lower curve is obtained by computing the dft within a Hamming window of duration 20 ms (200 samples). The upper curves on the graphs represent "pseudo-spectra". Each point is obtained by performing a weighted average of groups of points on the discrete Fourier transform over a bandwidth of 300 Hz. The three spectral sections are remarkably similar. Results are the same for the other two sets of data from this speaker. The other two speakers show the same degree of within-speaker consistency for this phenomenon.

Determining the onset and duration of the nasal murmurs in our entire data base turned out to be very difficult. In a study of error measurement magnitude, we discovered that measurement error was larger than the smallest value. We are currently trying to get around this. One technique involves biasing the measurements in favor of the statistical null hypothesis and applying special case analyses.

This problem itself seems at the heart of the language-specific differences in nasalization. At this stage, it appears that in English and

Figure 1

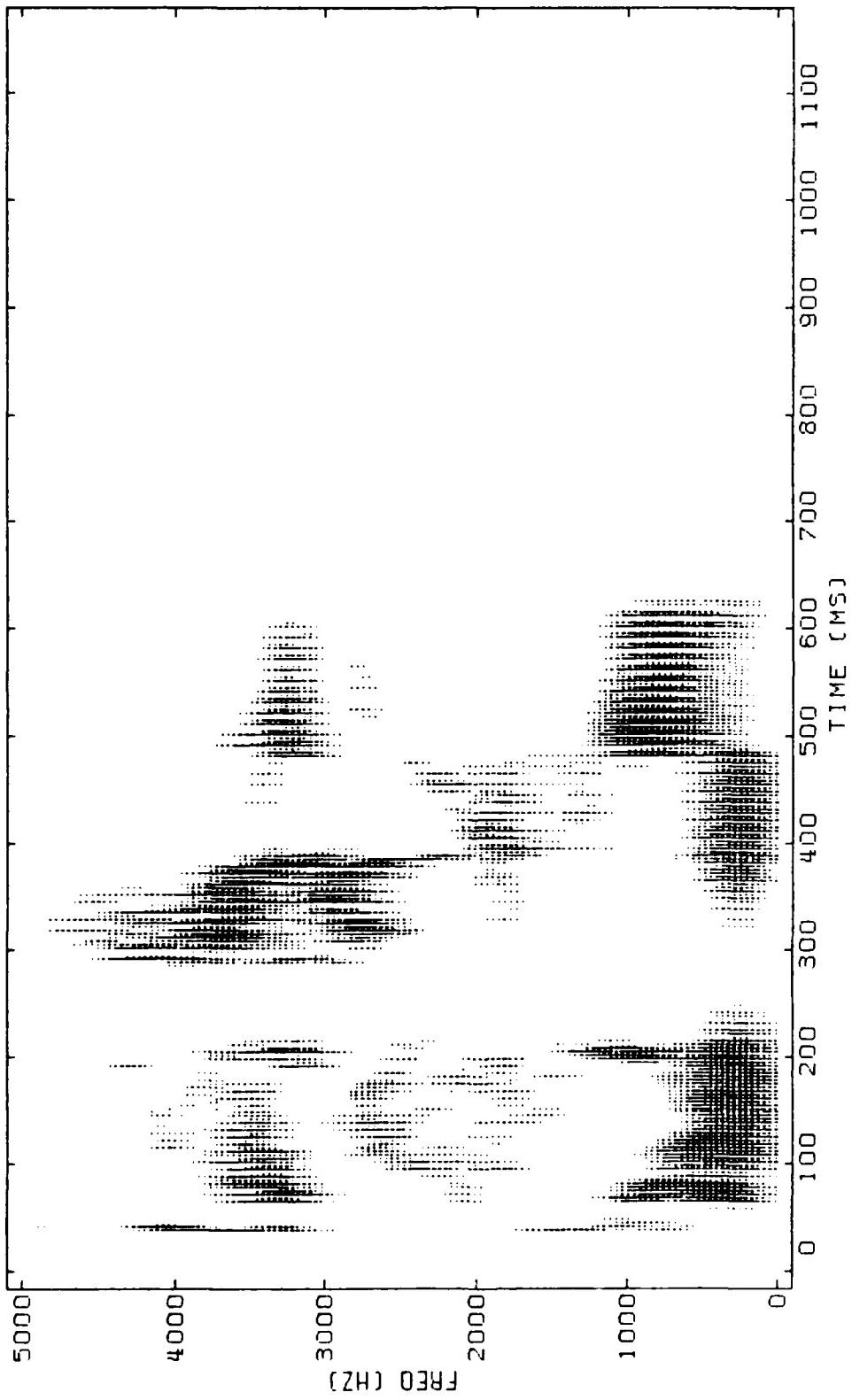


[LESLIE]

NOV 12 1986

FCDN12

condiment



D₂
E
.-
P
N
≈

Figure 2

Figure 3

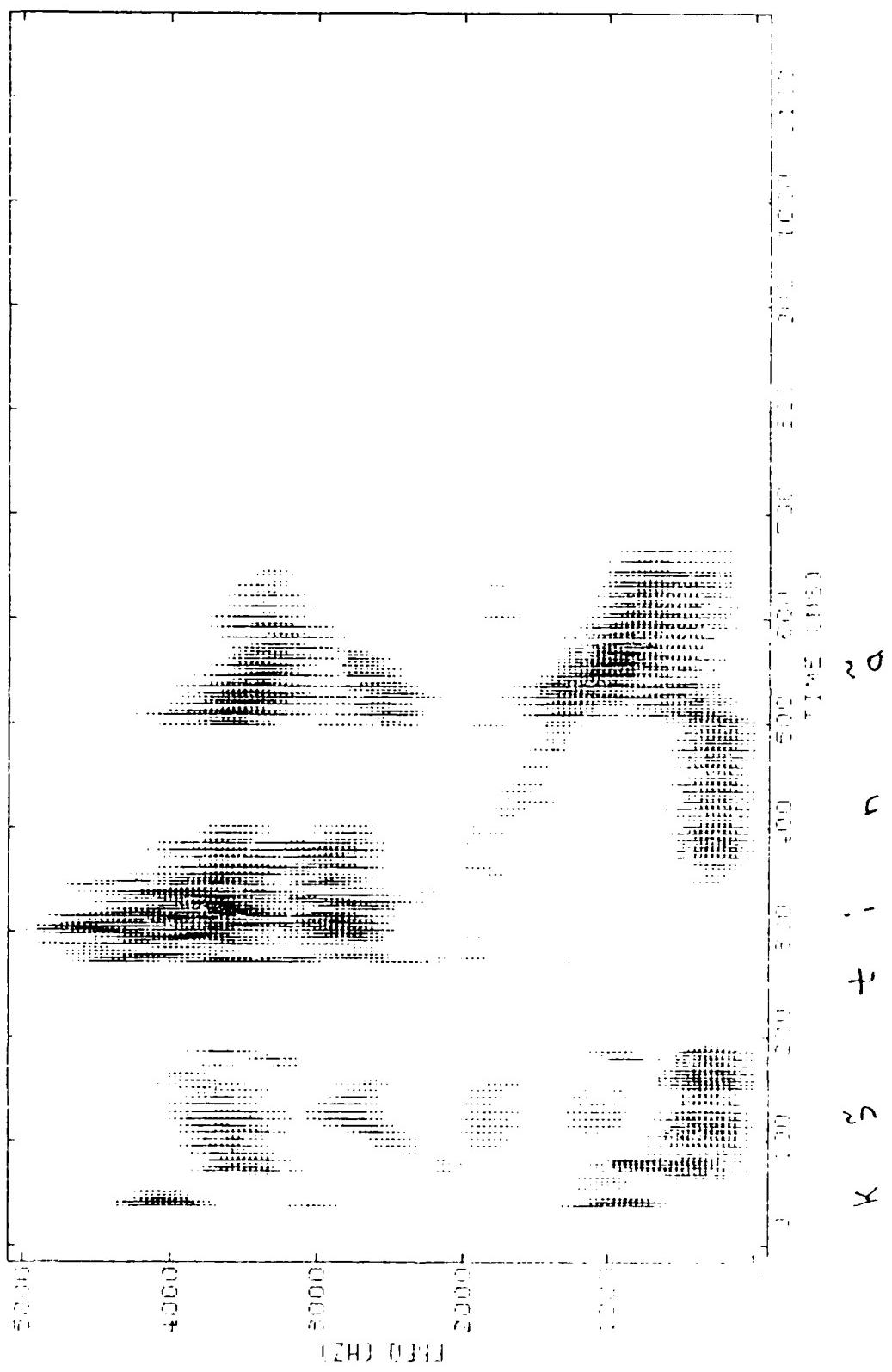


Figure 3

SEI 4 1986

LEEDY

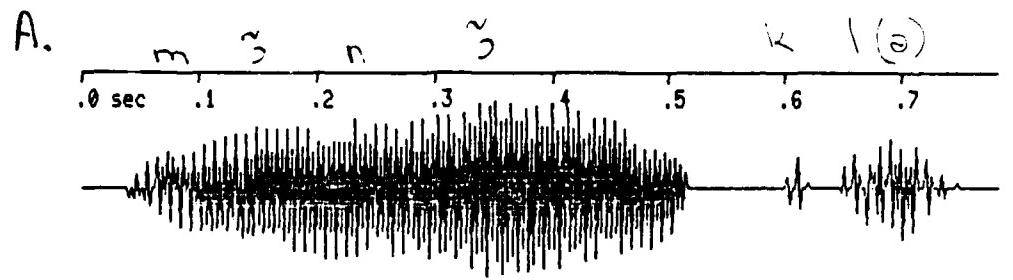


Figure 4

For each graph:

Above: amplitude envelope of the entire utterance.

Below: expanded waveform of 70 ms duration.

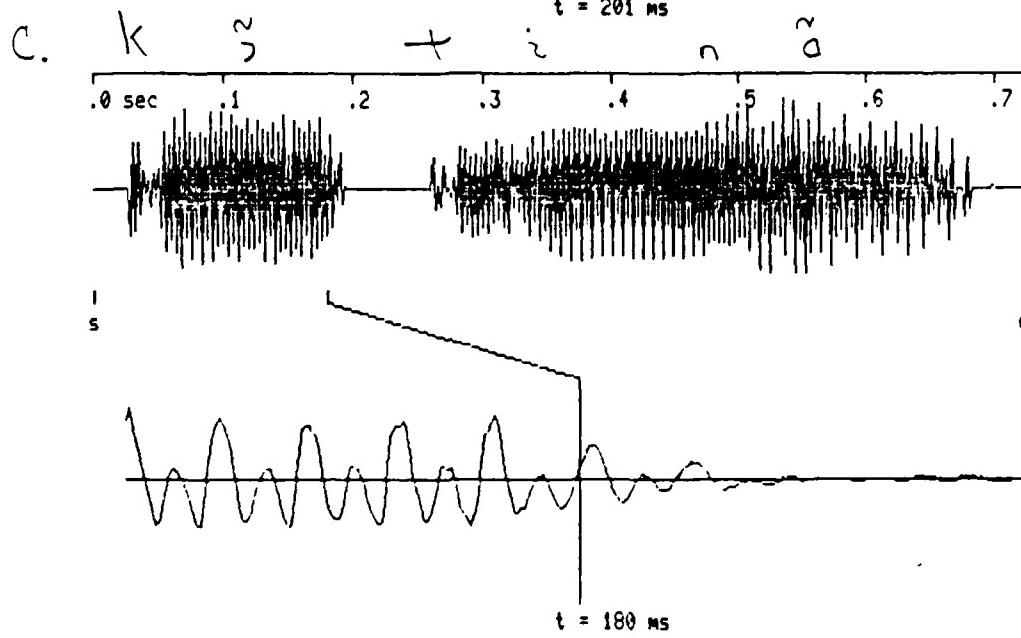
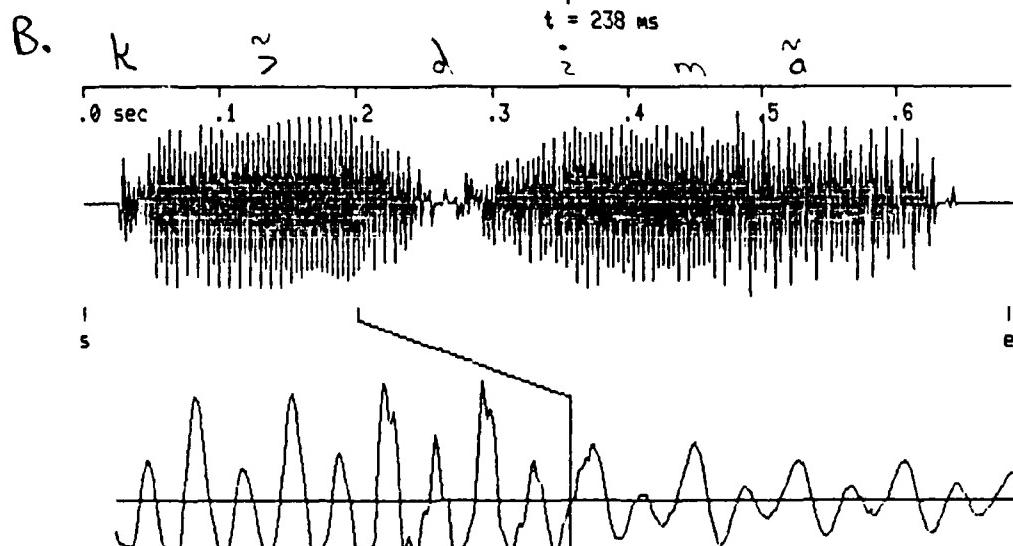
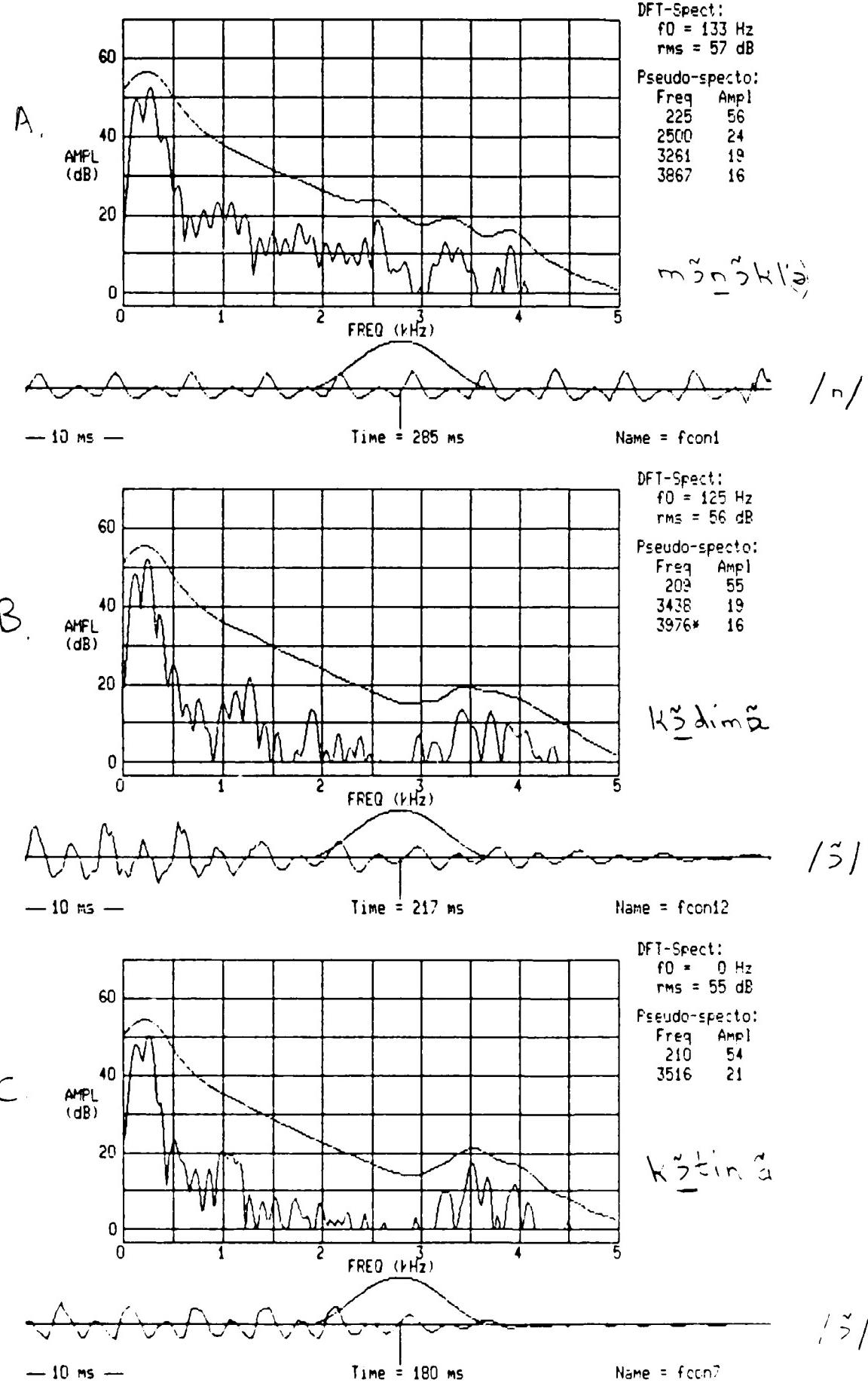


Figure 5



in Portuguese, the sequence of acoustic events occurring within utterances of these types is much easier to identify and measure than in French. This suggests some working hypotheses: In Portuguese and English, it is likely that (a) the lowering of the velum, the opening of the velar port, and oral tract closure in anticipation of the nasal and stop consonants covary temporally to such a degree that there is comparatively little vowel-dependent, consonant-dependent, or subject-dependent variation introduced, (b) Place of articulation for the nasal and following stop being the same, oral tract closure occurs very rapidly or abruptly so as to create a particularly distinct (and spectrographically visible) boundary between the vowel and the following consonants which emphasizes the onset of the first, the nasal consonant. In French it seems likely that (a) there is little or no temporal covariation between velar lowering / velar port opening and oral tract closure for the following consonant, but closure onset, rate, and duration are heavily dependent on the articulatory and acoustic features of the vowels themselves within which closure occurs. (b) oral tract closure in anticipation of the following stop occurs well in advance of the onset of the nasal murmur and is so gradual that the nasal murmur itself seems much more of a shortened anticlimactic transitional state between closure and release of the stop.

There are rather large differences in the timing and duration of events among subjects in the French study. We do not know whether this is an ideo-syncracy of the data base or whether these differences reflect greater variation in the French language itself, which would make boundaries less identifiable. For example, the spectral changes visible in the waveform just prior to the onset of the nasal murmurs in Figure 4 are ideoyncratic of oral tract closure for a following alveolar consonant and are particularly prominent in this vowel. In this subject, the spectral changes immediately precede a rather abrupt overall damping of energy in the waveform, and the durations of his nasal murmurs are, in general, short. For a second subject, however, these same spectral changes occur earlier proportionally in the vowel and are less prominent, loss of energy in the higher frequencies and overall damping of energy occur gradually, and the durations of his nasal murmurs are, in general, longer. In the first case, it is easy to identify murmur onsets, because the closeness in time of these changes creates a boundary visible in waveform, spectral sections and spectrographs. In the second case, it is very difficult to identify murmur onset, whatever the method used to do so, though the sequence of events is similar.

As a general rule of thumb for the set of utterances of all three subjects, the clearer the onset of the murmur, the short the murmur, and the more obscure the murmur onset the longer the murmur. The rule varies somewhat between voiced and unvoiced stop consonants, as voicing of the following stop seems to slow down all changes in the vowels, enhancing some and diminishing others.

Current and future work:

While attempting to resolve the problems involved in cross-language comparisons of durational characteristics, we ask ourselves: if the eye's ease in measurement is linked to covariation of measurement criteria, covariation whose degree and direction strengthen/weaken visible boundaries, what must the ear be doing and how different are the two?

thesized vowels by inserting a pole-zero pair in the vicinity of the first natural frequency in an all-pole transfer function. They varied the frequencies and spacing of the pole and zero to change the degree of nasalization and selected their parameter values on the basis of acoustic theory and feedback from prior experimentation. From their results they speculated that there is a distinctive change in the auditory response resulting from modification of the spectral prominence in the vicinity of the first resonant frequency. But, the psychoacoustic effects of these modifications (and those of other investigators) could not be explained only in terms of spectral measures. Many more questions arise:

- How does the auditory system respond to changes in the temporal patterning of the frequencies and spacing of the pole(s) and zero(s) of the transfer functions? And to changes in the "center of gravity" in vowel spectra?
- How does the auditory system respond, in general, to changes in the strength of the temporal covariation of spectral events? How close, or far apart, in time do these events need to occur to enhance or weaken the perception of acoustic boundaries? Does this depend on the nature of the events themselves?
- Are there cases in which strengthening negative covariation among events (i.e., increasing rate of change of one event and decreasing the rate of change of another) can strengthen the psychoacoustic impression of acoustic boundary?

We are currently attempting to quantify consistent changes, and limits of change, in the frequency, intensity and bandwidth of the spectral peaks below 2000 Hz for each of the French nasal vowel cases, and eventually for analogous English and Portuguese utterances, to characterize language-specific (and common) degrees and shapes of prominence and the direction and rate of their change. We hope to design continua of synthetic stimuli whose same parameters covary by progressively different degrees within the limits found in natural utterances. And we hope to introduce foreign accent into the synthesized utterances to study the language-specific variations in patterns of nasalization.

L.A. Wheeler

REFERENCES

1. Pickett, J. M., The Sounds of Speech Communication. University Park Press, Baltimore, 1980.
2. Fant, G., Acoustic Theory of Speech Production, Mouton and Co. 's-Gravenhage, Netherlands, 1960.
3. Ferguson, C.A., Hyman, L.M. and Ohala, J.J., Eds. Preface. In: Nasalfest. Papers from a Symposium on Nasals and Nasalization. Language Universals Project, Dept. Linguistics, Stanford University, Stanford, CA November, 1975.
4. Ferguson, C.A. Universal tendencies and 'normal' nasality. In; Nasalfest. Papers from a Symposium on Nasals and Nasalization. Language Universals Project, Dept. Linguistics, Stanford University, Stanford, CA, pp. 175-196. November, 1975.
5. Beddor, P.S., and Strange, W. Cross-language study of perception of the oral-nasal distinction, J. Acoust. Soc. Am. 71, pp. 1551-1561. 1982
6. Clumeck, H. A cross-linguistic investigation of vowel nasalization: an instrumental study. In: Nasalfest. Papers from a Symposium on Nasals and Nasalization. Language Universals Project, Dept. Linguistics, Stanford University, Stanford, Ca, pp.133-151. 1975
7. Foley, W.A. Some rules involving nasals and their implications. Ibid. pp. 213-230.
8. Malécot, A. Vowel nasality as a distinctive feature in American English, Language, 36, pp. 222-229.
9. Stevens, K.N., personal communications and pre-publication reports on collaborative studies with Amalia Andrade of the Center for Linguistics of the University of Lisbon, Portugal.
10. Hawkins, S., and Stevens, K.N. Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels. J. Acoust. Soc. Am., 77, pp. 1560-1575. 1985.

CLESLIE

DEC 4 1986

10000
50000
Mon 20000

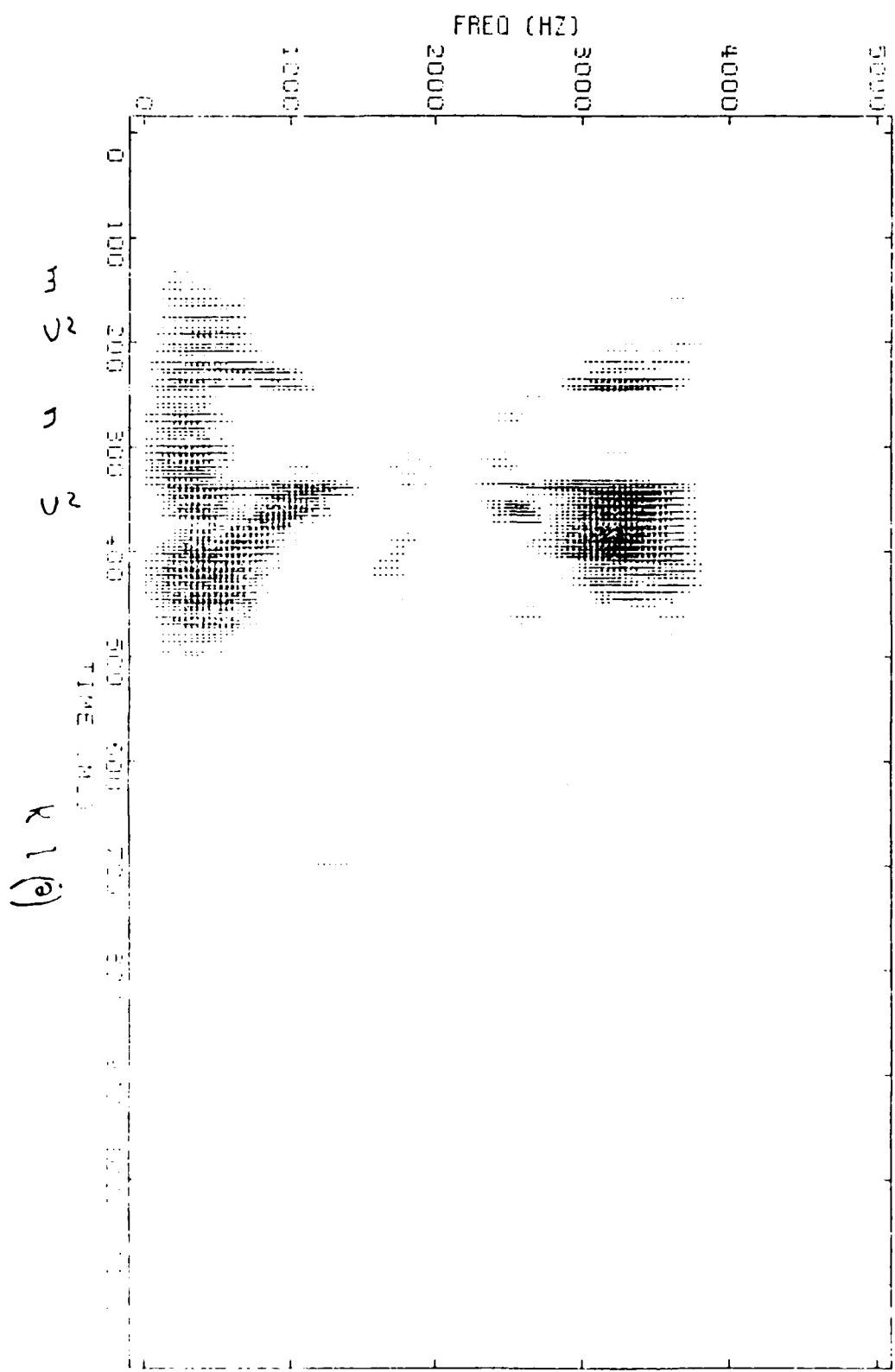


Figure 1

END

4 - 87

DTIC